BNL-91310-2010-CP

Domain Analysis of Computational Science

Fifty Years of a Scientific Computing group

Michiko Tanaka

Presented at the 11th International ISKO Conference
Rome, Italy
February 23-26, 2010

Information Services Division

Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
www.bnl.gov

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This preprint is intended for publication in a journal or proceedings. Since changes may be made before publication, it may not be cited or reproduced without the author's permission.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Printed on recycled paper

Michiko Tanaka
Brookhaven National Laboratory and Long Island University, Long Island, New York

Domain Analysis of Computational Science
Fifty Years of a Scientific Computing group

Abstract: I employed bibliometric- and historical-methods to study the domain of the Scientific Computing group at Brookhaven National Laboratory (BNL) for an extended period of fifty years, from 1958 to 2007. I noted and confirmed the growing emergence of interdisciplinarity within the group. I also identified a strong, consistent mathematics and physics orientation within it.

1. Introduction

At the core of domain analysis is the study and dissection of the activities and products of the domains to gain a sound understanding of their structures, their information needs and usages. With this valuable tool, information scientists acquire a detailed picture of the knowledge needed to support the work of a particular user group. I adopted this approach to study the Scientific Computing group at Brookhaven National Laboratory in the United States, following them over 50 years. This period began with the establishment of NASA in 1958, a year after the successful launch of the Soviet Union's Sputnik in 1957, and the year in which BNL started constructing the new high-capacity digital computer, later named Merlin. The 2007, the last year of my research period saw the installation at BNL of the new supercomputer Blue Gene/L, the current supercomputer, and the publication by NASA and 13 space agencies worldwide of their plans for space explorations.

Computational science, an interdisciplinary field in and of itself, emerged, as a discipline when supercomputing became the major focus of scientific computing. This fast-evolving area is founded upon collaborations and the sharing of ideas among all the group's members. However, a special issue arising from the domain's pluralism and metamorphism lies in its own intrinsic diversity and evolution. Therefore, a major problem facing information scientists is to identify the degree to which a computational science domain such as this is truly interdisciplinary, contributing to scientific knowledge externally as well as internally. The benefit of such an exploration will yield greater insights into the specific communication patterns used by computational scientists, and further advance the creation of knowledge across the scientific disciplines, while verifying the likely value of the approach to other complex scientific domains.

In this study, I attempted to refine the ambiguous definition of a group of scientists in a scientific computing group (1) through domain analysis, employing the well-demonstrated strength of bibliometric methods (Hjørland 2002, Tennis 2003), and, (2) gaining perspective by detailing the history of this group over half a century (1958-2007) (Hjørland 2002). My specific objectives were to identify 1) the characteristics of the computational scientists' scholarly communication; 2) their patterns of usage of scholarship to create and develop knowledge, and conduct interdisciplinary research; and, 3) the developmental history of computational science and the changes leading to its acceptance as a discrete domain.

2. Methodology and data analysis

2.1 Publication Trends

First, I identified the authors, detailed the publications of this Scientific Computing group, and recorded trends with time; in parallel, but to a more limited extent, I related these evolving orientations to national- and worldwide-events. The elements I amassed encompassed the numbers of publications, their type, such as journals versus reports, the number of authors for each publication, and the topic or classification of the journals and conference proceedings. I found that the expertise of the authors listed on the publications during the study period shifted from an initial primary focus on mathematics and physics to broader, interdisciplinary fields later. The number of authors on each publication also increased; in the first 25 years, an average of 62 percent of the documents had a single author whereas after 2003, only 3.5 percent of the publications were single-authored. This trend reflects the increasing complexity and interdisciplinarity of computational science. The number and type of publications that the group generated seemingly were influenced by the changes within BNL, in government policy, and the funding support for its research, and in major worldwide events. The launch of Sputnik 1 in 1957, the oil shortages in 1973, and recent concerns over national competitiveness, all generated new modes of governmental support of science and technology. Therefore, the research activities at National Laboratories, including BNL, significantly affected the development of computers and computational science (Seidel 1996). The number of publications of reports, many of which are keyed towards BNL's specific problems, such as its computers and its computing facility or networks, increased in the 1960s when they introduced CDC6600, which generally is considered as the first successful supercomputer for BNL's scientific community. Similarly, the group's publications, both in journals and conference papers, increased since 1999, when they became involved in SciDAC, the Department of Energy's program that called for developing computing at the level of 100 teraflops, and storing data in petabytes, and increasing their participation in scientific experiments in collaboration with other scientific departments.

2.2 Citation Analysis

I detailed the citation patterns of the group. I examined journals in which they published, and the 4859 cited references of those publications, to identify which authors and journals they themselves cite in their papers. About 29 percent (1405/4859) of the references are articles by the 50 most frequently cited authors. I consider these articles and books or book chapters as this domain's "citation classics" or "most influential" citations;they clearly revealed the group's mathematics orientation and interdisciplinarity characteristics. Philippe G. Ciarlet's 1978 book, *Finite Element Method,* a 1987 article by James H. Bramble, et al, in *Mathematical Computation,* and George A. Baker's 1965 article in *Advances in Theoretical Physics* are the top three most cited papers by the group.

There are 6514 citing articles for the group's publications. Citing journals symbolize the reach of this group's work; viz., its influence outside its own domain, the scope of which even from the beginning, extended into many diverse disciplines. Overall, my data validated the concept that Scientific Computing group is collaborative and interdisciplinary. The list exhibits diversified and interdisciplinary characteristics; it included such journals as *Analytical Chemistry, Journal of Magnetic Resonance Imaging, Proteomics, Astrophysics, Space Science,* and *Inorganic Chemistry.*

Further, from my exploration of publication patterns, a strong picture emerged of how a microcosmic scientific group evolved, grew, and matured to epitomize major national-

and international-events.

2.3 Author Cocitation Analysis

Contrary to cited and citing references, Author Cocitation Analysis (ACA) reveals indirect linkages or relations of authors and is invaluable for mapping and visualizing studies of co-citation (Small 1973, Small and Griffith 1974, McCain 1990, White and McCain 1989, Small 1999). I conducted ACA, using Thomson Reuters's Web of Science (WoS), based on the co-citation frequencies for twenty members of the Scientific Computing group. The bibliometric method of ACA uses authors as the unit of analysis and the co-citations of pairs of authors as the variable that indicates their distances from each other. The underlying assumption is that the more often pairs of authors are cited together, the closer is their academic relationship (White & Griffith 1981). The range of co-citations among 20 authors was 5 to 144; the highest mean co-citation count was 7.2. Using SPSS, I created an empirical map of the group's authors. The results of the cluster analysis include a Pearson correlation matrix and a dendogram showing the complete linkage results. I observed four clusters; the first and fourth fell into two sub-clusters.

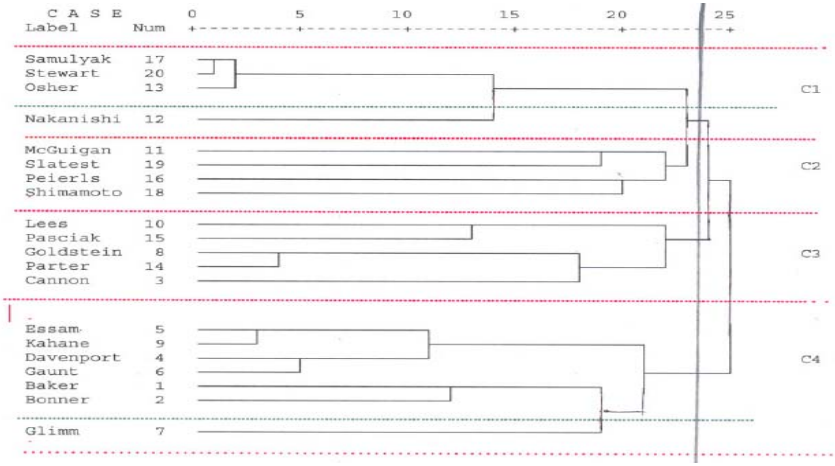

Figure 1. Dendogram using Complete Linkage

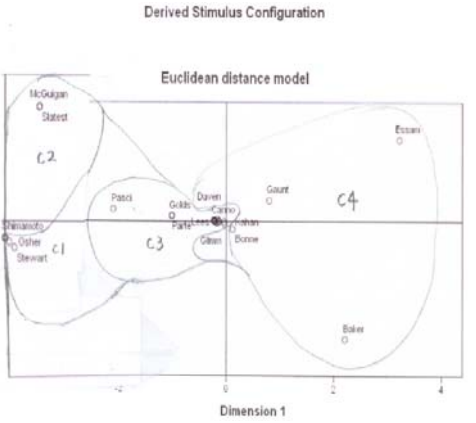

Figure 2. Clusters demonstrated

2.4 Content Analysis

Further, I explored the titles of the group's publications and analyzed them via WordStat; I verified and expanded my findings by identifying the terms used in them, from which I drew my conclusions on the group's interdisciplinarity. The content analysis of titles revealed the authors' specialization; their research clearly focused on the underlying mathematics and physics of computational science, though there was a notable increase in the frequency of interdisciplinary words in those titles published during the later years of my study period.

	19581962	19631967	19681972	1973197	19781982	19831987	19881992	19931997	19982002	20032007
BNL	0	5	12	12	2	0	1	2	6	4
CHEM	11	22	26	18	4	2	5	11	5	76
COM_SCI	14	114	136	114	52	22	38	51	52	144
INTERDIS	25	98	80	56	29	10	14	22	29	91
LIFE_SCI	0	6	4	1	0	0	3	5	2	38
MATH	42	153	242	121	85	57	58	56	32	157
PHYSICS	12	65	58	27	9	15	5	7	26	68

Table 1. Word Frequencies in the 10 periods

The titles of the cited references of the Scientific Computing group expanded my illustration of the groups' characteristics and research activities; they are mathematicians and physicists, primarily focused on these two areas. Together, references to these subjects constituted a little over 57 percent of all the title words I analyzed. I further examined titles of citing references, the content analysis of which are deemed to be valuable in reflecting the extension of this group, i.e., its axes of modulation (Tennis 2003, Smiraglia 2007). Compared with the content analysis of cited references of this group, these titles display even more interdisciplinarity; they encompass words in a

variety of disciplines, including ethanol, nanotube, nanostructure, and biomolecules. This finding reveals that the group's publications are widely read by scientists in many different fields, and that the influence of the group's researches on them is significant. The titles of Scientific Computing group's publications confirm its involvement in fields requiring computation, i.e., all of which Kenneth Wilson (1984) called "Grand Challenge"-level sciences.

3. Discussion

My study of computational science is unique for two reasons. First, I explored the domain microscopically, and second, I applied my findings macroscopically; thus, I used author co-citation and word frequencies of the titles of the group's publications for the former purpose (McCain 1983), and a historical study for the latter, which together afforded complementary views of the domain (McCain et al. 2005). From the emergence of computational science as a domain in the 1950s, I was able to follow its evolution throughout its entire history, from its genesis to its current state. I believe such a complete chronological coverage of a domain is singular, and that my methods are applicable to any new emerging domain, e.g., nanoscience. On the other hand, my study has remarkable similarities to others, supporting the value of this methodological approach. McCain's research on neural networks and software engineering, both new fields in applied science, are two examples. Comparable to her work, author co-citation analysis of the domain showed the relation between computational science and "foundational" work in the established disciplines, mathematics and physics. I observed the interdisciplinarity of the domain and links and their changes between computational science and the disciplines, i.e., the foundation research area (McCain 1998), that contributed to its development, and to changes as the field matured. Highly cited authors establish a cluster for a particular specialty (Small 1973). I noted that the specialty of researchers in this group lies in mathematics and physics with their underlying social relationships, as McCain (1983) indicated in her research on the macroeconomics domain, encompassing collaborations principally among themselves, to extending to other departments, and other institutions. This domain is a very dynamic one, as are other interdisciplinary domains. As I clearly demonstrated, researchers have explored and crossed boundaries of scientific disciplines; they even have chosen to publish their findings in the journals of other related domains, rather than in those of their own specialty. Conversely, researchers from other disciplines occasionally publish in the journals of the computational-science domain. Indeed, we see examples in the field of information science that researchers specializing in information retrieval are publishing more often in computer science journals (e.g., ACM journals) (Hjørland 2008).

My study, which revealed these phenomena of a powerful interdisciplinary domain, highlighted once again the value and efficacy of domain analysis in information science. As an iterative mathematical cornerstone of knowledge organization, it proved invaluable in allowing me to propose a socio-scientific structure for this Scientific Computing group.

4. Conclusions

Domain analysis and its approaches to studying domains are invaluable for gaining a detailed understanding of a particular domain. Bibliometrics, including ACA and content analysis of titles of cited and citing articles, helped me to identify two axes, the extension and intention of the domain, and confirmed this domain's interdisciplinarity. I further confirmed its interdisciplinarity by quantifying and visualizing the relations among

authors (ACA) and the title terms (content analysis). The comprehensive combination of these tools from information science, proved indispensable in supporting my characterization and delineation of the domain of computational science.

References

Hjørland, B., 2002, Domain analysis in information science: eleven approaches, *Journal of documentation* 58, 422-62.

Hjørland, B., 2008, Private communication on 8 August 2008.

McCain, K.W., 1983, The author co-citation structure of macroeconomics, *Scientometrics*, 5, 277-89.

McCain, K.W., 1990, Mapping authors in intellectual space: a technical overview, *Journal of the American society for information science*, 41, 433-43.

McCain, K.W., 1998, Neural networks research in context: a longitudinal JCA of an emerging interdisciplinary field, *Scientometrics*, 41, 389-410.

McCain, K.W., Verner, J.M., Hislop, G.W., et al., 2005, The use of bibliometric and knowledge elicitation techniques to map a knowledge domain: software engineering in the 1990s, *Scientometrics*, 65, 131-44.

Seidel, R.W., 1996, From Mars to Minverva: the origins of scientific computing in the AEC lab, *Physics today*, 49, n.10, 33-39.

Small, Henry. 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents, *Journal of the American society for information science, 24,* 799-813

Small, H., and Griffith., G., 1974, The structure of scientific literature I: Identifying and graphing specialties, *Science Studies*, 4, 17-40.

Small, H., 1999, Visualizing science by citation mapping, *Journal of the American society for information science*, 50, 799-813.

Smiraglia, R.P., 2007, Two kinds of power: insight into the legacy of Patrick Wilson in *Information sharingin a fragmented world: Crossing boundaries, Proceedings of the Canadian Association for Information Science*, *McGill University, Montréal, Quebec, 10 – 12 May, 2007,* ed. ClémentArsenault and Kimiz Dalkir, Available at <cais-acsi.ca/proceedings/2007/smiraglia_2007.pdf>

Tennis, J. T., 2003, Two axes of domains for domain analysis, *Knowledge organization,* 30, n.3/4,191-95.

White, H. D., and Griffith, B.C., 1981, Author co-citation: A literature measure of intellectual structure. *Journal of the American society for information science* 32, 163-72.

Wilson, K. G., 1987, Grand challenges to computational science, in *Modern physics in America: a Michelson-Morley centennial symposium. Cleveland, Ohio, 30-31 October 1987, AIP conference proceedings 169,* ed. William Fickinger and Kenneth L. Kowalski

White, H. D., and McCain, K., 1989, Bibliometrics, in *Annual Review of Information Science and Technology*, 24, 119-86.

Web documents have been accessed August 31, 2009.